

# A Sharp Condition for Exact Support Recovery With Orthogonal Matching Pursuit

Jinming Wen, Zhengchun Zhou, Jian Wang, Xiaohu Tang, and Qun Mo

**Abstract**—Support recovery of sparse signals from noisy measurements with orthogonal matching pursuit (OMP) has been extensively studied. In this paper, we show that for any  $K$ -sparse signal  $x$ , if a sensing matrix  $A$  satisfies the restricted isometry property (RIP) with restricted isometry constant (RIC)  $\delta_{K+1} < 1/\sqrt{K+1}$ , then under some constraints on the minimum magnitude of nonzero elements of  $x$ , OMP exactly recovers the support of  $x$  from its measurements  $y = Ax + v$  in  $K$  iterations, where  $v$  is a noise vector that is  $\ell_2$  or  $\ell_\infty$  bounded. This sufficient condition is sharp in terms of  $\delta_{K+1}$  since for any given positive integer  $K$  and any  $1/\sqrt{K+1} \leq \delta < 1$ , there always exists a matrix  $A$  satisfying the RIP with  $\delta_{K+1} = \delta$  for which OMP fails to recover a  $K$ -sparse signal  $x$  in  $K$  iterations. Also, our constraints on the minimum magnitude of nonzero elements of  $x$  are weaker than existing ones. Moreover, we propose worst-case necessary conditions for the exact support recovery of  $x$ , characterized by the minimum magnitude of the nonzero elements of  $x$ .

**Index Terms**—Compressed sensing (CS), restricted isometry property (RIP), restricted isometry constant (RIC), orthogonal matching pursuit (OMP), support recovery.

## I. INTRODUCTION

IN compressed sensing (CS), we frequently encounter the following linear model [1]–[4]:

$$y = Ax + v, \quad (1)$$

where  $x \in \mathbb{R}^n$  is an unknown  $K$ -sparse signal, (i.e.,  $|\text{supp}(x)| \leq K$ , where  $\text{supp}(x) = \{i : x_i \neq 0\}$  is the support of  $x$  and  $|\text{supp}(x)|$  is the cardinality of  $\text{supp}(x)$ ),  $A \in \mathbb{R}^{m \times n}$  ( $m \ll n$ ) is a known sensing matrix,  $y \in \mathbb{R}^m$  contains the noisy observations (measurements), and  $v \in \mathbb{R}^m$  is a noise vector. There are several common types of noises, such as the  $\ell_2$  bounded noise (i.e.,  $\|v\|_2 \leq \epsilon$  for some constant  $\epsilon$  [5]–[7]),

This work was supported by “Programme Avenir Lyon Saint-Etienne de l’Université de Lyon” in the framework of the programme “Investissements d’Avenir” (ANR-11-IDEX-0007), ANR through the HPAC project under Grant ANR 11 BS02 013, the NSF of China under grant 11271010 and 11531013, and the fundamental research funds for the Central Universities. This work was presented in part at the IEEE International Symposium on Information Theory (ISIT 2016), Barcelona, Spain.

J. Wen was with ENS de Lyon, Laboratoire LIP (U. Lyon, CNRS, ENSL, INRIA, UCBL), Lyon 69007, France. He is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton T6G 2V4, Canada (e-mail: jinming1@ualberta.ca).

Z. Zhou is with the School of Mathematics, Southwest Jiaotong University, Chengdu 610031, China, (e-mail: zzc@home.swjtu.edu.cn).

J. Wang is with the Department of Electrical and Computer Engineering, Duke University, USA, 27708, (e-mail: wangjianeee@gmail.com).

X. Tang is with the Information Security and National Computing Grid Laboratory, Southwest Jiaotong University, Chengdu 610031, China, (e-mail: xhutang@swjtu.edu.cn).

Q. Mo is with the Department of Mathematics, Zhejiang University, Hangzhou 310027, China (e-mail: moqun@zju.edu.cn).

---

### Algorithm 1 The OMP Algorithm [12]

---

Input:  $y$ ,  $A$ , and stopping rule.

Initialize:  $k = 0$ ,  $r^0 = y$ ,  $S_0 = \emptyset$ .

until the stopping rule is met

- 1:  $k = k + 1$ ,
- 2:  $s^k = \arg \max_{1 \leq i \leq n} |\langle r^{k-1}, A_i \rangle|$ ,
- 3:  $S_k = S_{k-1} \cup \{s^k\}$ ,
- 4:  $\hat{x}_{S_k} = \arg \min_{x \in \mathbb{R}^{|S_k|}} \|y - A_{S_k} x\|_2$ ,
- 5:  $r^k = y - A_{S_k} \hat{x}_{S_k}$ .

Output:  $\hat{x} = \arg \min_{x: \text{supp}(x) = S_k} \|y - Ax\|_2$ .

---

the  $\ell_\infty$  bounded noise (i.e.,  $\|Av\|_\infty \leq \epsilon$  for some constant  $\epsilon$  [8]), and the Gaussian noise (i.e.,  $v_i \sim \mathcal{N}(0, \sigma^2)$  [9]). In this paper, we consider only the first two types of noises, as the analysis for these two types can be easily extended to the last one by following some techniques in [8].

One of the central goals of CS is to recover the sparse signal  $x$  on the basis of the sensing matrix  $A$  and the observations  $y$ . It has been demonstrated that under appropriate conditions on  $A$ , the original signal  $x$  can be reliably recovered via properly designed algorithms [10] [11]. Orthogonal matching pursuit (OMP) [12] [13] is a widely-used greedy algorithm for performing the recovery task. For any set  $S \subset \{1, 2, \dots, n\}$ , let  $A_S$  denote the submatrix of  $A$  that contains only the columns indexed by  $S$ . Similarly, let  $x_S$  denote the subvector of  $x$  that contains only the entries indexed by  $S$ . Then, the OMP algorithm is formally described in Algorithm 1.<sup>1</sup>

A widely used framework for analyzing the recovery performance of the CS recovery algorithms is the restricted isometry property (RIP) [1]. For an  $m \times n$  matrix  $A$  and any integer  $K$ , the order- $K$  restricted isometry constant (RIC)  $\delta_K$  is defined as the smallest constant such that

$$(1 - \delta_K) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_K) \|x\|_2^2 \quad (2)$$

for all  $K$ -sparse vectors  $x$ .

For the noise-free case (i.e., when  $v = 0$ ), many RIC-based conditions have been proposed to guarantee the exact recovery of sparse signals via OMP. It has respectively been shown in [14] and [15] that  $\delta_{K+1} < 1/(3\sqrt{K})$  and  $\delta_{K+1} < 1/\sqrt{2K}$  are sufficient for OMP to recover any  $K$ -sparse signal  $x$  in  $K$  iterations. Later, the conditions have been improved to  $\delta_{K+1} < 1/(1 + \sqrt{K})$  [16] [17] and further to  $\delta_{K+1} < (\sqrt{4K+1} -$

<sup>1</sup>If the maximum correlation in Step 2 occurs for multiple indices, break the tie randomly.

1)/(2K) [18]. Recently, it has been shown that if  $\delta_{K+1} < 1/\sqrt{K+1}$ , OMP is guaranteed to exactly recover  $K$ -sparse signals  $\mathbf{x}$  in  $K$  iterations [19]. On the other hand, it has been conjectured in [20] that there exists a matrix  $\mathbf{A}$  satisfying the RIP with  $\delta_{K+1} \leq 1/\sqrt{K}$  such that OMP fails to recover a  $K$ -sparse vector  $\mathbf{x}$  in  $K$  iterations. This conjecture has been confirmed by examples provided in [16] [17]. Furthermore, it has been reported in [21] [19] that for any given positive integer  $K \geq 2$  and any given  $\delta$  satisfying  $1/\sqrt{K+1} \leq \delta < 1$ , there always exist a  $K$ -sparse vector  $\mathbf{x}$  and a matrix  $\mathbf{A}$  satisfying the RIP with  $\delta_{K+1} = \delta$  such that the OMP algorithm fails to recover  $\mathbf{x}$  in  $K$  iterations. In other words, sufficient conditions for recovering  $\mathbf{x}$  with  $K$  steps of OMP cannot be weaker than  $\delta_{K+1} < 1/\sqrt{K+1}$ , which therefore implies that  $\delta_{K+1} < 1/\sqrt{K+1}$  is a sharp condition [19].

For the noisy case (i.e., when  $\mathbf{v} \neq \mathbf{0}$ ), we are often interested in recovering the support of  $\mathbf{x}$ , i.e.,  $\text{supp}(\mathbf{x})$ . Once  $\text{supp}(\mathbf{x})$  is exactly recovered, the underlying signal  $\mathbf{x}$  can be easily estimated by ordinary least squares regression [8]. It has been shown in [22] that under some constraint on the minimum magnitude of nonzero elements of  $\mathbf{x}$  (i.e.,  $\min_{i \in \text{supp}(\mathbf{x})} |x_i|$ ),  $\delta_{K+1} < 1/(\sqrt{K} + 3)$  is sufficient for OMP to exactly recover  $\text{supp}(\mathbf{x})$  under both the  $\ell_2$  and  $\ell_\infty$  bounded noises. The sufficient condition has been improved to  $\delta_{K+1} < 1/(\sqrt{K} + 1)$  [23], and the best existing condition in terms of  $\delta_{K+1}$  is  $\delta_{K+1} < (\sqrt{4K+1} - 1)/(2K)$  [18].

In this paper, we investigate sufficient, and *worst-case* necessary conditions, based on the RIC and  $\min_{i \in \text{supp}(\mathbf{x})} |x_i|$ , for recovering  $\text{supp}(\mathbf{x})$  with OMP under both  $\ell_2$  and  $\ell_\infty$  bounded noises. Here, the worst-case necessity means that if it is violated, then there is (at least) one instance of  $\{\mathbf{A}, \mathbf{x}, \mathbf{v}\}$  such that OMP fails to recover  $\text{supp}(\mathbf{x})$  from the noisy measurements  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  [24]. Specifically, our contributions can be summarized as follows.

- i) We show that if  $\mathbf{A}$  and  $\mathbf{v}$  in (1) respectively satisfy the RIP with  $\delta_{K+1} < 1/\sqrt{K+1}$  and  $\|\mathbf{v}\|_2 \leq \epsilon$ , then OMP with the stopping rule  $\|\mathbf{r}^k\|_2 \leq \epsilon$  exactly recovers  $\text{supp}(\mathbf{x})$  in  $K$  iterations, provided that

$$\min_{i \in \text{supp}(\mathbf{x})} |x_i| > \frac{2\epsilon}{1 - \sqrt{K+1}\delta_{K+1}}.$$

We also show that our constraint on  $\min_{i \in \text{supp}(\mathbf{x})} |x_i|$  is weaker than existing ones.<sup>2</sup> (Theorem 1).

- ii) We show that if  $\mathbf{A}$  and  $\mathbf{v}$  in (1) respectively satisfy the RIP with  $\delta_{K+1} < 1/\sqrt{K+1}$  and  $\|\mathbf{A}^T \mathbf{v}\|_\infty \leq \epsilon$ , then OMP with the stopping rule

$$\|\mathbf{A}^T \mathbf{r}^k\|_\infty \leq \left(1 + \sqrt{\frac{(1+\delta_2)K}{1-\delta_{K+1}}}\right) \epsilon$$

exactly recovers  $\text{supp}(\mathbf{x})$  in  $K$  iterations, provided that

$$\min_{i \in \text{supp}(\mathbf{x})} |x_i| > \frac{2}{1 - \sqrt{K+1}\delta_{K+1}} \left(1 + \sqrt{\frac{(1+\delta_2)K}{1-\delta_{K+1}}}\right) \epsilon.$$

We also compare our constraint on  $\min_{i \in \text{supp}(\mathbf{x})} |x_i|$  with existing results (Theorem 2).

- iii) We show that for any given positive integer  $K$ ,  $0 < \delta < 1/\sqrt{K+1}$ , and  $\epsilon > 0$ , there always exist a sensing matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  satisfying the RIP with  $\delta_{K+1} = \delta$ , a  $K$ -sparse vector  $\mathbf{x} \in \mathbb{R}^n$  with

$$\min_{i \in \text{supp}(\mathbf{x})} |x_i| < \frac{\sqrt{1-\delta}\epsilon}{1 - \sqrt{K+1}\delta},$$

and a noise vector  $\mathbf{v} \in \mathbb{R}^m$  with  $\|\mathbf{v}\|_2 \leq \epsilon$ , such that OMP fails to recover  $\text{supp}(\mathbf{x})$  from  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  in  $K$  iterations (Theorem 3).

- iv) We show that for any given positive integer  $K$ ,  $0 < \delta < 1/\sqrt{K+1}$  and  $\epsilon > 0$ , there always exist a sensing matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  satisfying the RIP with  $\delta_{K+1} = \delta$ , a  $K$ -sparse vector  $\mathbf{x} \in \mathbb{R}^n$  with

$$\min_{i \in \text{supp}(\mathbf{x})} |x_i| < \frac{2\epsilon}{1 - \sqrt{K+1}\delta},$$

and a noise vector  $\mathbf{v} \in \mathbb{R}^m$  with  $\|\mathbf{A}^T \mathbf{v}\|_\infty \leq \epsilon$ , such that OMP fails to recover  $\text{supp}(\mathbf{x})$  from  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  in  $K$  iterations (Theorem 4).

Since OMP may fail to recover a  $K$ -sparse signal with OMP in  $K$  iterations when  $\mathbf{A}$  satisfies the RIP with  $\delta \geq 1/\sqrt{K+1}$  and  $\mathbf{v} = \mathbf{0}$  [21] [19], sufficient conditions for recovering  $\text{supp}(\mathbf{x})$  with OMP in  $K$  iterations in the noisy case cannot be weaker than  $\delta < 1/\sqrt{K+1}$  (note that  $\mathbf{v} = \mathbf{0}$  is the ideal case). Hence, our sufficient conditions summarized in i)–ii) are sharp in terms of the RIC. Moreover, iii) and iv) indicate that for all  $K$ -sparse vectors  $\mathbf{x}$  and sensing matrices  $\mathbf{A}$  satisfying the RIP with  $\delta_{K+1} < 1/\sqrt{K+1}$ , the worst-case necessary constraint on  $\min_{i \in \text{supp}(\mathbf{x})} |x_i|$  guaranteeing the exact recovery of  $\text{supp}(\mathbf{x})$  from (1) with  $K$  iterations of OMP are

$$\min_{i \in \text{supp}(\mathbf{x})} |x_i| \geq \frac{\sqrt{1-\delta_{K+1}}\epsilon}{1 - \sqrt{K+1}\delta_{K+1}},$$

and

$$\min_{i \in \text{supp}(\mathbf{x})} |x_i| \geq \frac{2\epsilon}{1 - \sqrt{K+1}\delta_{K+1}},$$

under the  $\ell_2$  and  $\ell_\infty$  bounded noise, respectively.

The rest of the paper is organized as follows. In Section II, we introduce some notations that will be used throughout this paper. We also propose a lemma which plays a central role in proving our new sufficient conditions. In Section III, we present sufficient, and worst-case necessary conditions for the exact support recovery of sparse signals with OMP under both the  $\ell_2$  and  $\ell_\infty$  bounded noises. Finally, we conclude our paper in Section IV.

## II. PRELIMINARIES

### A. Notation

We first define some notations that will be used throughout this paper. Let  $\mathbb{R}$  be the real field. Boldface lowercase letters denote column vectors, and boldface uppercase letters denote matrices. For a vector  $\mathbf{x}$ ,  $\mathbf{x}_{i:j}$  denotes the subvector of  $\mathbf{x}$  formed by entries  $i, i+1, \dots, j$ . Let  $\mathbf{e}_k$  denote the  $k$ -th column of the identity matrix  $\mathbf{I}$  and  $\mathbf{0}$  denote a zero matrix or a zero column vector. Denote  $\Omega = \text{supp}(\mathbf{x})$  and  $|\Omega|$  be the cardinality of  $\Omega$ , then for any  $K$ -sparse signal  $\mathbf{x}$ ,  $|\Omega| \leq K$ .

<sup>2</sup>These results were presented at the 2016 IEEE International Symposium on Information Theory (ISIT) conference [25].

For any set  $S$ , denote  $\Omega \setminus S = \{i | i \in \Omega, i \notin S\}$ . Let  $\Omega^c$  and  $S^c$  denote the complement of  $\Omega$  and  $S$ , respectively, i.e.,  $\Omega^c = \{1, 2, \dots, n\} \setminus \Omega$ , and  $S^c = \{1, 2, \dots, n\} \setminus S$ . Let  $\mathbf{A}_S$  denote the submatrix of  $\mathbf{A}$  that contains only the columns indexed by  $S$ . Similarly, let  $\mathbf{x}_S$  denote the subvector of  $\mathbf{x}$  that contains only the entries indexed by  $S$ . For any matrix  $\mathbf{A}_S$  of full column-rank, let  $\mathbf{P}_S = \mathbf{A}_S(\mathbf{A}_S^T \mathbf{A}_S)^{-1} \mathbf{A}_S^T$  and  $\mathbf{P}_S^\perp = \mathbf{I} - \mathbf{P}_S$  denote the projector and orthogonal complement projector onto the column space of  $\mathbf{A}_S$ , respectively, where  $\mathbf{A}_S^T$  stands for the transpose of  $\mathbf{A}_S$ .

### B. A useful lemma

We present the following lemma, which is one of the central results of this paper and will play a key role in proving our sufficient conditions for support recovery with OMP.

*Lemma 1:* Suppose that  $\mathbf{A}$  in (1) satisfies the RIP of order  $K+1$  and  $S$  is a proper subset of  $\Omega$  (i.e.,  $S \subset \Omega$  with  $|S| < |\Omega|$ ). Then,

$$\begin{aligned} & \|\mathbf{A}_{\Omega \setminus S}^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}\|_\infty - \|\mathbf{A}_{\Omega^c}^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}\|_\infty \\ & \geq \frac{(1 - \sqrt{|\Omega| - |S| + 1} \delta_{|\Omega|+1}) \|\mathbf{x}_{\Omega \setminus S}\|_2}{\sqrt{|\Omega| - |S|}}. \end{aligned} \quad (3)$$

*Proof.* See Appendix A.  $\square$

Note that in the noise-free case (i.e., when  $\mathbf{v} = \mathbf{0}$ ), Lemma 1 can be directly connected to the selection rule of OMP. Specifically, if we assume that  $S = S_k \subset \Omega$  for some  $0 \leq k < |\Omega|$  (see Algorithm 1 for the definition of  $S_k$ ), then

$$\mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S} = \mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k} = \mathbf{P}_{S_k}^\perp \mathbf{A} \mathbf{x} = \mathbf{P}_{S_k}^\perp \mathbf{y} = \mathbf{r}^k \quad (4)$$

and (3) can be rewritten as

$$\begin{aligned} & \max_{j \in \Omega \setminus S_k} |\langle \mathbf{A}_j, \mathbf{r}^k \rangle| - \max_{j \in \Omega^c} |\langle \mathbf{A}_j, \mathbf{r}^k \rangle| \\ & \geq \frac{(1 - \sqrt{|\Omega| - k + 1} \delta_{|\Omega|+1}) \|\mathbf{x}_{\Omega \setminus S_k}\|_2}{\sqrt{|\Omega| - k}}. \end{aligned} \quad (5)$$

Clearly, (5) characterizes a lower bound on the difference between the maximum value of the OMP decision-metric for the columns belonging to  $\Omega$  and that for the columns belonging to  $\Omega^c$ . Since  $\|\mathbf{x}_{\Omega \setminus S_k}\|_2 > 0$  for  $k < |\Omega|$ , (5) implies that OMP chooses a correct index among  $\Omega$  in the  $(k+1)$ -th iteration as long as  $\delta < 1/\sqrt{|\Omega| - k + 1}$ . Thus, by induction, one can show that OMP exactly recovers  $\Omega$  in  $K$  iterations under  $\delta < 1/\sqrt{K+1}$ , which matches the result in [19].

In the noisy case (i.e., when  $\mathbf{v} \neq \mathbf{0}$ ), by assuming that  $S = S_k \subset \Omega$  for some  $0 \leq k < |\Omega|$ , we have

$$\mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S} = \mathbf{P}_{S_k}^\perp \mathbf{A} \mathbf{x} = \mathbf{P}_{S_k}^\perp \mathbf{y} - \mathbf{P}_{S_k}^\perp \mathbf{v} = \mathbf{r}^k - \mathbf{P}_{S_k}^\perp \mathbf{v}. \quad (6)$$

Due to the extra term  $\mathbf{P}_{S_k}^\perp \mathbf{v}$  in (6), however, we cannot directly obtain (5) from (3). Nevertheless, by applying (6) (i.e., the relationship between  $\mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}$  and  $\mathbf{r}^k$ ), one can implicitly obtain from (3) a lower bound for

$$\max_{j \in \Omega \setminus S_k} |\langle \mathbf{A}_j, \mathbf{r}^k \rangle| - \max_{j \in \Omega^c} |\langle \mathbf{A}_j, \mathbf{r}^k \rangle|$$

by utilizing  $\|\mathbf{x}_{\Omega \setminus S_k}\|_2$ ,  $\delta_{|\Omega|+1}$  and  $\mathbf{v}$ . The lower bound, together with some constraints on  $\min_{i \in \Omega} |x_i|$ , allows to build

sufficient conditions for OMP to select an index belonging to  $\Omega$  at the  $(k+1)$ -th iteration. See more details in Section III-A.

*Remark 1:* Lemma 1 is motivated by [19, Lemma II.2]. But these two lemmas have a key distinction. Specifically, while [19, Lemmas II.2] showed that

$$\|\mathbf{A}_{\Omega}^T \mathbf{A} \mathbf{x}\|_\infty > \|\mathbf{A}_{\Omega^c}^T \mathbf{A} \mathbf{x}\|_\infty,$$

Lemma 1 quantitatively characterizes a lower bound on

$$\|\mathbf{A}_{\Omega \setminus S}^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}\|_\infty - \|\mathbf{A}_{\Omega^c}^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}\|_\infty,$$

which is stronger in the following aspects:

- i) Lemma 1 is more general and embraces [19, Lemma II.2] as a special case. To be specific, Lemma 1 holds for any  $S$  which is a proper subset of  $\Omega$ , while [19, Lemma II.2] only works for the case where  $S = \emptyset$ . In fact, the generality of Lemma 1 (i.e., it works for any  $S \subset \Omega$ ) is of vital importance for the noisy case analysis of OMP. Indeed, due to the noise involved, the recovery condition for the first iteration of OMP does not apply to the succeeding iterations.<sup>3</sup> Thus, we need to consider the recovery condition for every individual iteration of OMP, which, as will be seen later, essentially corresponds to the cases of  $S = S_k$ ,  $k = 0, 1, \dots, K-1$ , in Lemma 1.
- ii) In contrast to [19, Lemma II.2] that is applicable to the noise-free case only, the lower bound in Lemma 1 works for both the noise-free as well as the noisy case (as indicated above). Specifically, as will be seen in Appendix B, by applying this quantitative lower bound with  $S = S_k$ , together with the relationship between the residual and  $\mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k}$  (see (6)), we are able to get a precise characterization of the difference between the maximum value of OMP decision metrics for correct and incorrect columns in the noisy case, from which the sufficient condition guaranteeing a correct selection immediately follows.
- iii) Compared to [19, Lemma II.2], Lemma 1 gives a sharper lower bound on the difference between the maximum value of the OMP decision-metric for the columns belonging to  $\Omega$  and that for the columns belonging to  $\Omega^c$ . Specifically, [19, Lemma II.2] showed that  $\|\mathbf{A}_{\Omega}^T \mathbf{A} \mathbf{x}\|_\infty - \|\mathbf{A}_{\Omega^c}^T \mathbf{A} \mathbf{x}\|_\infty$  is lower bounded by zero when  $\mathbf{A}$  satisfies the RIP with  $\delta_{K+1} < 1/\sqrt{K+1}$ . Whereas, applying  $S = \emptyset$  in (3) yields

$$\|\mathbf{A}_{\Omega}^T \mathbf{A} \mathbf{x}\|_\infty - \|\mathbf{A}_{\Omega^c}^T \mathbf{A} \mathbf{x}\|_\infty \geq (1 - \sqrt{|\Omega| + 1} \delta_{|\Omega|+1}) \min_{i \in \Omega} |x_i|,$$

where the right-hand side can be much larger than zero under the same RIP assumption.

## III. MAIN ANALYSIS

In this section, we will show that if a sensing matrix  $\mathbf{A}$  satisfies the RIP with  $\delta_{K+1} < 1/\sqrt{K+1}$ , then under some constraints on  $\min_{i \in \Omega} |x_i|$ , OMP exactly recovers  $\text{supp}(\mathbf{x})$  in  $K$  iterations from the noisy measurements  $\mathbf{y} = \mathbf{A} \mathbf{x} + \mathbf{v}$ .

<sup>3</sup>This is in contrast to the noise-free case where the condition for the first iteration of OMP immediately applies to the succeeding iterations because the residual of those iterations can be viewed as the *modified measurements* of  $K$ -sparse signals with the same sensing matrix  $\mathbf{A}$ ; See, e.g., [17, Lemma 6].

We will also present worst-case necessary conditions on  $\min_{i \in \Omega} |x_i|$  for the exact recovery of  $\text{supp}(\mathbf{x})$ .

#### A. Sufficient condition

We consider both  $\ell_2$  and  $\ell_\infty$  bounded noises. The following theorem gives a sufficient condition for the exact support recovery with OMP under the  $\ell_2$  bounded noise.

*Theorem 1:* Suppose that  $\mathbf{A}$  and  $\mathbf{v}$  in (1) respectively satisfy the RIP with

$$\delta_{K+1} < \frac{1}{\sqrt{K+1}}, \quad (7)$$

and  $\|\mathbf{v}\|_2 \leq \epsilon$ . Then OMP with the stopping rule  $\|\mathbf{r}^k\|_2 \leq \epsilon$  exactly recovers the support  $\Omega$  of any  $K$ -sparse signal  $\mathbf{x}$  from  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  in  $|\Omega|$  iterations, provided that

$$\min_{i \in \Omega} |x_i| > \frac{2\epsilon}{1 - \sqrt{K+1}\delta_{K+1}}. \quad (8)$$

*Proof.* See Appendix B.  $\square$

If  $\mathbf{v} = \mathbf{0}$ , then we can set  $\epsilon = 0$  which implies that (8) holds. Thus we have the following result, which coincides with [19, Theorem III.1].

*Corollary 1:* If  $\mathbf{A}$  and  $\mathbf{v}$  in (1) respectively satisfy the RIP with (7) and  $\mathbf{v} = \mathbf{0}$ , then OMP exactly recovers all  $K$ -sparse signals  $\mathbf{x}$  from  $\mathbf{y} = \mathbf{A}\mathbf{x}$  in  $K$  iterations.

*Remark 2:* In [21] and [19], it has been shown that for any given integer  $K \geq 2$  and for any  $1/\sqrt{K+1} \leq \delta < 1$ , there always exist a  $K$ -sparse vector  $\mathbf{x}$  and a sensing matrix  $\mathbf{A}$  with  $\delta_{K+1} = \delta$ , such that the OMP algorithm fails to recover  $\mathbf{x}$  from (1) (note that this statement also holds for  $K = 1$ ). Therefore, the sufficient condition in Theorem 1 for the exact support recovery with OMP is sharp in terms of  $\delta_{K+1}$ .

It might be interesting to compare our condition with existing results. In [18], [22], [23], similar recovery conditions have been proposed for the OMP algorithm under the assumption that the sensing matrices  $\mathbf{A}$  have normalized columns (i.e.,  $\|\mathbf{A}_i\|_2 = 1$  for  $i = 1, 2, \dots, n$ ). Comparing with these conditions, our condition given by Theorem 1 is more general as it works for sensing matrices whose column  $\ell_2$ -norms are not necessarily equal to 1. More importantly, our result is less restrictive than those in [18], [22], [23] with respect to both  $\delta_{K+1}$  and  $\min_{i \in \Omega} |x_i|$ . To illustrate this, we compare our condition with that in [18], which is the best result to date. In [18], it was shown that if the sensing matrix  $\mathbf{A}$  is column normalized and satisfies the RIP with

$$\delta_{K+1} < \frac{\sqrt{4K+1} - 1}{2K}$$

and the noise vector  $\mathbf{v}$  satisfies  $\|\mathbf{v}\|_2 \leq \epsilon$ , then the OMP algorithm with the stopping rule  $\|\mathbf{r}^k\|_2 \leq \epsilon$  exactly recovers the support  $\Omega$  of any  $K$ -sparse signal  $\mathbf{x}$  from  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  in  $K$  iterations, provided that

$$\min_{i \in \Omega} |x_i| > \frac{(\sqrt{1 + \delta_{K+1}} + 1)\epsilon}{1 - \delta_{K+1} - \sqrt{1 - \delta_{K+1}}\sqrt{K}\delta_{K+1}}.$$

To show our condition in Theorem 1 is less restrictive, it suffices to show that

$$\frac{\sqrt{4K+1} - 1}{2K} < \frac{1}{\sqrt{K+1}} \quad (9)$$

and that

$$\frac{(\sqrt{1 + \delta_{K+1}} + 1)\epsilon}{1 - \delta_{K+1} - \sqrt{1 - \delta_{K+1}}\sqrt{K}\delta_{K+1}} > \frac{2\epsilon}{1 - \sqrt{K+1}\delta_{K+1}} \quad (10)$$

To show (9), we need to show

$$\sqrt{(4K+1)(K+1)} < 2K + \sqrt{K+1},$$

which is equivalent to

$$4K^2 + 5K + 1 < 4K^2 + K + 1 + 4K\sqrt{K+1}.$$

Since  $K \geq 1$ , the aforementioned equation holds, so (9) holds.

We next focus on the proof of (10). Since  $\sqrt{1 + \delta_{K+1}} + 1 > 2$ , it is clear that (10) holds if

$$1 - \delta_{K+1} - \sqrt{1 - \delta_{K+1}}\sqrt{K}\delta_{K+1} < 1 - \sqrt{K+1}\delta_{K+1}.$$

Equivalently,

$$1 + \sqrt{1 - \delta_{K+1}}\sqrt{K} > \sqrt{K+1}. \quad (11)$$

Obviously, (11) holds if

$$\sqrt{1 - \delta_{K+1}} > \frac{\sqrt{K+1} - 1}{\sqrt{K}},$$

which is equivalent to

$$\delta_{K+1} < \frac{2(\sqrt{K+1} - 1)}{K}.$$

By (7), it suffices to show

$$\frac{1}{\sqrt{K+1}} < \frac{2(\sqrt{K+1} - 1)}{K}.$$

By some simple calculations, one can show that the aforementioned inequality holds. Thus, (10) holds under (7).

Now we turn to the case where the noise vector  $\mathbf{v}$  is  $\ell_\infty$  bounded.

*Theorem 2:* Suppose that  $\mathbf{A}$  and  $\mathbf{v}$  in (1) respectively satisfy the RIP with (7) and  $\|\mathbf{A}^T \mathbf{v}\|_\infty \leq \epsilon$ . Then OMP with the stopping rule

$$\|\mathbf{A}^T \mathbf{r}^k\|_\infty \leq \left(1 + \sqrt{\frac{1 + \delta_2}{1 - \delta_{K+1}}}\sqrt{K}\right)\epsilon \quad (12)$$

exactly recovers the support  $\Omega$  of any  $K$ -sparse signal  $\mathbf{x}$  from  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  in  $|\Omega|$  iterations, provided that<sup>4</sup>

$$\min_{i \in \Omega} |x_i| > \frac{2}{1 - \sqrt{K+1}\delta_{K+1}} \left(1 + \sqrt{\frac{1 + \delta_2}{1 - \delta_{K+1}}}\sqrt{K}\right)\epsilon. \quad (13)$$

*Proof.* See Appendix C.  $\square$

*Remark 3:* While in [18], [22], [23],

$$\|\mathbf{A}^T \mathbf{r}^k\|_\infty \leq \epsilon \quad (14)$$

was used as the stopping rule of OMP, we would like to note that (12) in Theorem 2 cannot be replaced by (14). Otherwise, OMP may choose indices that do not belong to

<sup>4</sup>If the columns of  $\mathbf{A}$  exhibit a unit  $\ell_2$  norm, then (12) and (13) can be respectively relaxed to  $\|\mathbf{A}^T \mathbf{r}^k\|_\infty \leq \left(1 + \frac{\sqrt{K}}{\sqrt{1 - \delta_{K+1}}}\right)\epsilon$  and  $\min_{i \in \Omega} |x_i| > \frac{2}{1 - \sqrt{K+1}\delta_{K+1}} \left(1 + \frac{\sqrt{K}}{\sqrt{1 - \delta_{K+1}}}\right)\epsilon$ .

$\Omega$ , no matter how large  $\min_{i \in \Omega} |x_i|$  is. To illustrate this, we give an example, where for simplicity we consider sensing matrix  $\mathbf{A}$  with  $\ell_2$ -normalized columns.

*Example 1:* For any  $0 < \delta < 1$  and any  $a > \frac{1+\delta}{1-\delta}$ , let

$$\mathbf{A} = \begin{bmatrix} \sqrt{1-\delta^2} & 0 \\ \delta & 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} a \\ 0 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -\frac{2\delta}{\sqrt{1-\delta^2}} \\ 1 \end{bmatrix}, \epsilon = 1.$$

Then,  $\mathbf{x}$  is an 1-sparse vector supported on  $\Omega = \{1\}$  and

$$\min_{i \in \Omega} |x_i| = a.$$

$\mathbf{A}$  has unit  $\ell_2$ -norm columns. It is easy to verify that the singular values of  $\mathbf{A}^T \mathbf{A}$  are  $1 \pm \delta$ , which, by the definition of the RIC, implies that  $\delta_2 = \delta$ . Moreover, since

$$\mathbf{A}^T \mathbf{v} = \begin{bmatrix} -\delta \\ 1 \end{bmatrix},$$

we have  $\|\mathbf{A}^T \mathbf{v}\|_\infty \leq \epsilon$ . In the following, we show that if (14) is used as the stopping rule, then OMP finally returns the index set  $\{1, 2\}$ , no matter how large  $a$  is.

By (1), we have

$$\mathbf{y} = \begin{bmatrix} a\sqrt{1-\delta^2} - \frac{2\delta}{\sqrt{1-\delta^2}} \\ a\delta + 1 \end{bmatrix}.$$

Thus,

$$\mathbf{A}_1^T \mathbf{y} = a - \delta, \mathbf{A}_2^T \mathbf{y} = a\delta + 1.$$

Since  $a > \frac{1+\delta}{1-\delta}$ , we have  $\mathbf{A}_1^T \mathbf{y} > \mathbf{A}_2^T \mathbf{y}$ . Thus by the selection rule of OMP (see Algorithm 1),  $S_1 = \{1\}$  is identified and consequently,

$$\hat{x}_{S_1} = (\mathbf{A}_1^T \mathbf{A}_1)^{-1} \mathbf{A}_1^T \mathbf{y} = a - \delta$$

and the residual is updated as

$$\mathbf{r}^1 = \mathbf{y} - \mathbf{A}_1 \hat{x}_{S_1} = \begin{bmatrix} \delta\sqrt{1-\delta^2} - \frac{2\delta}{\sqrt{1-\delta^2}} \\ 1 + \delta^2 \end{bmatrix}.$$

By some calculations, we obtain

$$\mathbf{A}^T \mathbf{r}^1 = \begin{bmatrix} 0 \\ 1 + \delta^2 \end{bmatrix},$$

which implies that  $\|\mathbf{A}^T \mathbf{r}^1\|_\infty = 1 + \delta^2 > \epsilon$  so that the stopping condition (14) does not satisfy. Hence, the OMP algorithm will continue to the second iteration and will eventually return the index set  $\{1, 2\}$ .

Again, by [21] and [19], the sufficient condition given in Theorem 2 is sharp in terms of  $\delta_{K+1}$ . We mention that similar constraints on  $\min_{i \in \Omega} |x_i|$  have been proposed in [18], [22], [23]. However, since those results were based on a different stopping rule (i.e., (14)), we do not give a comparison of our constraint to those results.

### B. Worst-case necessary condition

In the above subsection, we have presented sufficient conditions guaranteeing exact support recovery of sparse signals with OMP. In this subsection, we investigate worst-case necessary conditions for the exact support recovery. Like the sufficient conditions, our necessary conditions are given in terms of the RIC as well as the minimum magnitude of

the nonzero elements of input signals. Note that OMP may fail to recover a  $K$ -sparse signal  $\mathbf{x}$  from  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  if  $\delta_{K+1} \geq 1/\sqrt{K+1}$ , even in the noise-free case [21] [19]. Hence,  $\delta_{K+1} < 1/\sqrt{K+1}$  naturally becomes a necessity for the noisy case. Therefore, in deriving the worst-case necessary condition on  $\min_{i \in \Omega} |x_i|$ , we consider only the matrices  $\mathbf{A}$  satisfying the RIP with  $\delta_{K+1} < 1/\sqrt{K+1}$ .

We first restrict our attention to the case of  $\ell_2$  bounded noises.

*Theorem 3:* For any given  $\epsilon > 0$ , positive integer  $K$ , and

$$0 < \delta < \frac{1}{\sqrt{K+1}}, \quad (15)$$

there always exist a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  satisfying the RIP with  $\delta_{K+1} = \delta$ , a  $K$ -sparse vector  $\mathbf{x} \in \mathbb{R}^n$  with

$$\min_{i \in \Omega} |x_i| < \frac{\sqrt{1-\delta}\epsilon}{1-\sqrt{K+1}\delta},$$

and a noise vector  $\mathbf{v} \in \mathbb{R}^m$  with  $\|\mathbf{v}\|_2 \leq \epsilon$ , such that OMP fails to recover  $\Omega$  from  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  in  $K$  iterations.

*Proof.* See Appendix D.  $\square$

*Remark 4:* One can immediately obtain from Theorem 3 that under the  $\ell_2$  bounded noise, a worst-case necessary condition (recall that the worst-case necessity means that if it is violated, then there is (at least) one instance of  $\{\mathbf{A}, \mathbf{x}, \mathbf{v}\}$  such that OMP fails to recover  $\text{supp}(\mathbf{x})$  from the noisy measurements  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  [24].) on  $\min_{i \in \Omega} |x_i|$  for OMP is:

$$\min_{i \in \Omega} |x_i| \geq \frac{\sqrt{1-\delta_{K+1}}\epsilon}{1-\sqrt{K+1}\delta_{K+1}}. \quad (16)$$

Here, we would like to note that the *worst-case* necessity does not mean that for  $\mathbf{A}$ ,  $\mathbf{x}$  and  $\mathbf{v}$ , (16) has to be satisfied to ensure the exact support recovery. In fact, it can be shown that OMP may be able to recover  $\text{supp}(\mathbf{x})$  in  $K$  iterations when (16) does not hold. One such example is given as follows.

*Example 2:* For any given  $0 < \delta < 1/\sqrt{2}$  and

$$0 < a < \frac{\sqrt{2(1-\delta)}}{1-\sqrt{K+1}\delta},$$

let

$$\mathbf{A} = \delta \mathbf{I}_2, \quad \mathbf{x} = \begin{bmatrix} a \\ 0 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \epsilon = \sqrt{2}.$$

Then,  $\mathbf{A}$  satisfies the RIP with  $\delta_2 = \delta < 1/\sqrt{2}$ . Moreover,  $\mathbf{x}$  is 1-sparse and does not satisfy (16). However, one can check that OMP with the stopping condition  $\|\mathbf{r}^k\|_2 \leq \epsilon$  exactly recovers  $\Omega$  in just one iteration.

*Remark 5:* We mention that the worst-case necessary condition for the exact support recovery with OMP has also been studied in [26], in which the author characterized the worst-case necessity using the signal-to-noise ratio (SNR). However, their result concerned only the sensing matrices  $\mathbf{A}$  whose singular values are ones. In comparison, our condition is more general and works for all sensing matrices  $\mathbf{A}$  satisfying the RIP with  $\delta_{K+1} < 1/\sqrt{K+1}$ .

Next, we proceed to the worst-case necessity analysis for the case where the noise is  $\ell_\infty$  bounded.

*Theorem 4:* For any given  $\epsilon > 0$ , positive integer  $K$ , and  $\delta$  satisfying (15), there always exist a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$

satisfying the RIP with  $\delta_{K+1} = \delta$ , a  $K$ -sparse vector  $\mathbf{x} \in \mathbb{R}^n$  with

$$\min_{i \in \Omega} |x_i| < \frac{2\epsilon}{1 - \sqrt{K+1}\delta},$$

and a noise vector  $\mathbf{v} \in \mathbb{R}^m$  with  $\|\mathbf{A}^T \mathbf{v}\|_\infty \leq \epsilon$ , such that OMP fails to recover  $\Omega$  from  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  in  $K$  iterations.

*Proof.* See Appendix E.  $\square$

*Remark 6:* Similar to the case of the  $\ell_2$  bounded noise, Theorem 4 implies a worst-case necessary condition on  $\min_{i \in \Omega} |x_i|$ , for exactly recovering  $\text{supp}(\mathbf{x})$  with OMP under the  $\ell_\infty$  bounded noise is:

$$\min_{i \in \Omega} |x_i| \geq \frac{2\epsilon}{1 - \sqrt{K+1}\delta_{K+1}}. \quad (17)$$

Again, we note that (17) applies to the worst case. For general cases, however, OMP may be able to exactly recover  $\Omega$  without this requirement. See a toy example as follows.

*Example 3:* For any given  $0 < \delta < 1/\sqrt{2}$  and

$$0 < a < \frac{2\sqrt{2}\delta}{1 - \sqrt{K+1}\delta},$$

let

$$\mathbf{A} = \delta \mathbf{I}_2, \quad \mathbf{x} = \begin{bmatrix} a \\ 0 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \epsilon = \sqrt{2}\delta.$$

Then,  $\mathbf{A}$  satisfies the RIP with  $\delta_2 = \delta < 1/\sqrt{2}$ .  $\mathbf{x}$  is 1-sparse and does not satisfy (17). Furthermore, one can easily show that the OMP algorithm can exactly recover  $\Omega$  in one iteration when the stopping rule in (12) is used.

Finally, we would like to mention that while our sufficient conditions are sharp in terms of the RIC, there is still some gap between the sufficient and the worst-case necessary constraints on  $\min_{i \in \text{supp}(\mathbf{x})} |x_i|$ . In particular, for the  $\ell_2$  bounded noise, the gap between conditions (8) and (16) is relatively small, demonstrating the tightness of the sufficient condition (8). For the  $\ell_\infty$  bounded noise, however, the gap between conditions (13) and (17) can be large since the expression  $\left(1 + \sqrt{\frac{1+\delta_2}{1-\delta_{K+1}}} \sqrt{K}\right)$  on the right-hand side of (13) can be much larger than one for a support cardinality  $K$  that is large enough. Whether it is possible to bridge this gap is an interesting open question.

#### IV. CONCLUSION

In this paper, we have studied sufficient conditions for the exact support recovery of sparse signals from noisy measurements by OMP. For both  $\ell_2$  and  $\ell_\infty$  bounded noises, we have shown that if the RIC of a sensing matrix  $\mathbf{A}$  satisfies the RIP with  $\delta_{K+1} < 1/\sqrt{K+1}$ , then under some conditions on the minimum magnitude of nonzero elements of the  $K$ -sparse signal  $\mathbf{x}$ , the support of  $\mathbf{x}$  can be exactly recovered in  $K$  iterations of OMP. The proposed conditions are sharp in terms of  $\delta_{K+1}$  for both types of noises, and the conditions on the minimum magnitude of nonzero elements of  $\mathbf{x}$  are weaker than existing ones. We have also proposed worst-case necessary conditions for the exact support recovery of  $\mathbf{x}$  characterized by the minimum magnitude of the nonzero elements of  $\mathbf{x}$ , under both  $\ell_2$  and  $\ell_\infty$  bounded noises.

#### ACKNOWLEDGMENT

We are grateful to the editor and the anonymous referees for their valuable and thoughtful suggestions that greatly improve the quality of this paper.

#### APPENDIX A PROOF OF LEMMA 1

Before proving Lemma 1, we introduce the following three useful lemmas, which were respectively proposed in [1], [27] and [28].

*Lemma 2:* If  $\mathbf{A}$  satisfies the RIP of orders  $k_1$  and  $k_2$  with  $k_1 < k_2$ , then  $\delta_{k_1} \leq \delta_{k_2}$ .

*Lemma 3:* Let  $\mathbf{A} \in \mathbb{R}^{m \times n}$  satisfy the RIP of order  $k$  and  $S \subset \{1, 2, \dots, n\}$  with  $|S| \leq k$ , then for any  $\mathbf{x} \in \mathbb{R}^m$ ,

$$\|\mathbf{A}_S^T \mathbf{x}\|_2^2 \leq (1 + \delta_k) \|\mathbf{x}\|_2^2.$$

*Lemma 4:* Let sets  $S_1, S_2$  satisfy  $|S_2 \setminus S_1| \geq 1$  and matrix  $\mathbf{A}$  satisfy the RIP of order  $|S_1 \cup S_2|$ , then for any vector  $\mathbf{x} \in \mathbb{R}^{|S_2 \setminus S_1|}$ ,

$$(1 - \delta_{|S_1 \cup S_2|}) \|\mathbf{x}\|_2^2 \leq \|\mathbf{P}_{S_1}^\perp \mathbf{A}_{S_2 \setminus S_1} \mathbf{x}\|_2^2 \leq (1 + \delta_{|S_1 \cup S_2|}) \|\mathbf{x}\|_2^2.$$

*Proof of Lemma 1.* Obviously, to show (3), it suffices to show for each  $j \in \Omega^c$ ,

$$\begin{aligned} & \|\mathbf{A}_{\Omega \setminus S}^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}\|_\infty - |\mathbf{A}_j^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}| \\ & \geq \frac{(1 - \sqrt{|\Omega| - |S| + 1} \delta_{|\Omega|+1}) \|\mathbf{x}_{\Omega \setminus S}\|_2}{\sqrt{|\Omega| - |S|}}. \end{aligned} \quad (18)$$

Since  $S$  is a proper subset of  $\Omega$ ,  $\|\mathbf{x}_{\Omega \setminus S}\|_1 \neq 0$ . Hence

$$\begin{aligned} & \|\mathbf{A}_{\Omega \setminus S}^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}\|_\infty \\ & = \frac{\sum_{\ell \in \Omega \setminus S} |x_\ell|}{\|\mathbf{x}_{\Omega \setminus S}\|_1} \|\mathbf{A}_{\Omega \setminus S}^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}\|_\infty \\ & \stackrel{(a)}{\geq} \frac{1}{\sqrt{|\Omega| - |S|} \|\mathbf{x}_{\Omega \setminus S}\|_2} \left( \sum_{\ell \in \Omega \setminus S} |x_\ell| \right) \|\mathbf{A}_{\Omega \setminus S}^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}\|_\infty \\ & \geq \frac{1}{\sqrt{|\Omega| - |S|} \|\mathbf{x}_{\Omega \setminus S}\|_2} \sum_{\ell \in \Omega \setminus S} (x_\ell \mathbf{A}_\ell^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}) \\ & = \frac{1}{\sqrt{|\Omega| - |S|} \|\mathbf{x}_{\Omega \setminus S}\|_2} \left( \sum_{\ell \in \Omega \setminus S} x_\ell \mathbf{A}_\ell \right)^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S} \\ & = \frac{1}{\sqrt{|\Omega| - |S|} \|\mathbf{x}_{\Omega \setminus S}\|_2} (\mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S})^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S} \\ & \stackrel{(b)}{=} \frac{1}{\sqrt{|\Omega| - |S|} \|\mathbf{x}_{\Omega \setminus S}\|_2} (\mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S})^T \mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S} \\ & = \frac{1}{\sqrt{|\Omega| - |S|} \|\mathbf{x}_{\Omega \setminus S}\|_2} \|\mathbf{P}_S^\perp \mathbf{A}_{\Omega \setminus S} \mathbf{x}_{\Omega \setminus S}\|_2^2, \end{aligned}$$

where (a) follows from  $|\text{supp}(\mathbf{x}_{\Omega \setminus S})| = |\Omega| - |S|$  and the fact that  $\|\mathbf{x}\|_1 \leq \sqrt{|\text{supp}(\mathbf{x})|} \|\mathbf{x}\|_2$  for all  $\mathbf{x} \in \mathbb{R}^n$  (For more details, see, e.g., [29, p.517]. Note that this inequality itself can be derived from the Cauchy-Schwarz inequality), and (b) is because  $\mathbf{P}_S^\perp$  is an orthogonal projector which has the idempotent and symmetry properties, i.e.,

$$(\mathbf{P}_S^\perp)^T \mathbf{P}_S^\perp = \mathbf{P}_S^\perp \mathbf{P}_S^\perp = \mathbf{P}_S^\perp. \quad (19)$$

Thus,

$$\begin{aligned} & \|P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}\|_2^2 \\ & \leq \sqrt{|\Omega| - |S|} \|x_{\Omega \setminus S}\|_2 \|A_{\Omega \setminus S}^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}\|_\infty. \end{aligned} \quad (20)$$

Let

$$\alpha = -\frac{\sqrt{|\Omega| - |S| + 1} - 1}{\sqrt{|\Omega| - |S|}}. \quad (21)$$

Then, by some simple calculations, we obtain

$$\frac{2\alpha}{1 - \alpha^2} = -\sqrt{|\Omega| - |S|}, \quad \frac{1 + \alpha^2}{1 - \alpha^2} = \sqrt{|\Omega| - |S| + 1}. \quad (22)$$

To simplify the notation, for given  $j \in \Omega^c$ , we define

$$B = P_S^\perp [A_{\Omega \setminus S} \quad A_j], \quad (23)$$

$$u = [x_{\Omega \setminus S}^T \quad 0]^T \in \mathbb{R}^{|\Omega \setminus S| + 1},$$

$$w = [0^T \quad \alpha t \|x_{\Omega \setminus S}\|_2]^T \in \mathbb{R}^{|\Omega \setminus S| + 1}, \quad (24)$$

where

$$t = \begin{cases} 1 & \text{if } A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S} \geq 0 \\ -1 & \text{if } A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S} < 0 \end{cases}. \quad (25)$$

Then,

$$P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S} = Bu, \quad (26)$$

and

$$\|u + w\|_2^2 = (1 + \alpha^2) \|x_{\Omega \setminus S}\|_2^2, \quad (27)$$

$$\|\alpha^2 u - w\|_2^2 = \alpha^2 (1 + \alpha^2) \|x_{\Omega \setminus S}\|_2^2. \quad (28)$$

Thus

$$\begin{aligned} w^T B^T B u & \stackrel{(a)}{=} \alpha t \|x_{\Omega \setminus S}\|_2 A_j^T (P_S^\perp)^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S} \\ & \stackrel{(b)}{=} \alpha t \|x_{\Omega \setminus S}\|_2 A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S} \\ & \stackrel{(c)}{=} \alpha \|x_{\Omega \setminus S}\|_2 |A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}|, \end{aligned}$$

where (a) follows from (23), (24) and (26), (b) follows from (19), and (c) is from (25). Therefore, for  $j \in \Omega^c$ , we have

$$\begin{aligned} & \|B(u + w)\|_2^2 \\ & = \|Bu\|_2^2 + \|Bw\|_2^2 + 2w^T B^T B u \\ & = \|Bu\|_2^2 + \|Bw\|_2^2 + 2\alpha \|x_{\Omega \setminus S}\|_2 |A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}| \end{aligned}$$

and

$$\begin{aligned} & \|B(\alpha^2 u - w)\|_2^2 \\ & = \alpha^4 \|Bu\|_2^2 + \|Bw\|_2^2 - 2\alpha^3 \|x_{\Omega \setminus S}\|_2 |A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}|. \end{aligned}$$

By the aforementioned equations, we have

$$\begin{aligned} & \|B(u + w)\|_2^2 - \|B(\alpha^2 u - w)\|_2^2 \\ & = (1 - \alpha^4) \|Bu\|_2^2 \\ & \quad + 2\alpha(1 + \alpha^2) \|x_{\Omega \setminus S}\|_2 |A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}| \\ & = (1 - \alpha^4) \left( \|Bu\|_2^2 + \frac{2\alpha}{1 - \alpha^2} \|x_{\Omega \setminus S}\|_2 |A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}| \right) \\ & = (1 - \alpha^4) \\ & \quad \times (\|Bu\|_2^2 - \sqrt{|\Omega| - |S|} \|x_{\Omega \setminus S}\|_2 |A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}|), \end{aligned} \quad (29)$$

where the last equality follows from the first equality in (22).

It is not hard to check that

$$\begin{aligned} & \|B(u + w)\|_2^2 - \|B(\alpha^2 u - w)\|_2^2 \\ & \stackrel{(a)}{\geq} (1 - \delta_{|\Omega|+1}) \|(u + w)\|_2^2 - (1 + \delta_{|\Omega|+1}) \|(\alpha^2 u - w)\|_2^2 \\ & \stackrel{(b)}{=} (1 - \delta_{|\Omega|+1}) (1 + \alpha^2) \|x_{\Omega \setminus S}\|_2^2 \\ & \quad - (1 + \delta_{|\Omega|+1}) \alpha^2 (1 + \alpha^2) \|x_{\Omega \setminus S}\|_2^2 \\ & = (1 + \alpha^2) \|x_{\Omega \setminus S}\|_2^2 ((1 - \delta_{|\Omega|+1}) - (1 + \delta_{|\Omega|+1}) \alpha^2) \\ & = (1 + \alpha^2) \|x_{\Omega \setminus S}\|_2^2 ((1 - \alpha^2) - \delta_{|\Omega|+1} (1 + \alpha^2)) \\ & = (1 - \alpha^4) \|x_{\Omega \setminus S}\|_2^2 \left( 1 - \frac{1 + \alpha^2}{1 - \alpha^2} \delta_{|\Omega|+1} \right) \\ & \stackrel{(c)}{=} (1 - \alpha^4) \|x_{\Omega \setminus S}\|_2^2 (1 - \sqrt{|\Omega| - |S| + 1} \delta_{|\Omega|+1}), \end{aligned} \quad (30)$$

where (a) follows from Lemma 4 and (23), (b) is from (27) and (28), and (c) follows from the second equality in (22).

By (26), (29), (30) and the fact that  $1 - \alpha^4 > 0$ , we have

$$\begin{aligned} & \|P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}\|_2^2 - \sqrt{|\Omega| - |S|} \|x_{\Omega \setminus S}\|_2 |A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}| \\ & = \|Bu\|_2^2 - \sqrt{|\Omega| - |S|} \|x_{\Omega \setminus S}\|_2 |A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}| \\ & \geq \|x_{\Omega \setminus S}\|_2^2 (1 - \sqrt{|\Omega| - |S| + 1} \delta_{|\Omega|+1}). \end{aligned}$$

Combining the aforementioned equation with (20), we obtain

$$\begin{aligned} & \sqrt{|\Omega| - |S|} \|x_{\Omega \setminus S}\|_2 \\ & \quad \times (\|A_{\Omega \setminus S}^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}\|_\infty - |A_j^T P_S^\perp A_{\Omega \setminus S} x_{\Omega \setminus S}|) \\ & \geq \|x_{\Omega \setminus S}\|_2^2 (1 - \sqrt{|\Omega| - |S| + 1} \delta_{|\Omega|+1}). \end{aligned}$$

Therefore, (18) holds, which establishes the lemma.  $\square$

## APPENDIX B PROOF OF THEOREM 1

*Proof.* Our proof consists of two steps. First, we show that OMP makes correct selection at each iteration. Then, we prove that it performs exactly  $|\Omega|$  iterations.

We prove the first step by induction. Suppose that the OMP algorithm selects correct indices in the first  $k-1$  iterations, i.e.,  $S_{k-1} \subseteq \Omega$ . Then, we need to show that the OMP algorithm also selects a correct index at the  $k$ -th iteration, i.e., showing that  $s^k \in \Omega$  (see Algorithm 1). Here, we assume  $1 \leq k \leq |\Omega|$ . Thus, the proof for the first selection corresponds to the case of  $k = 1$ . Clearly the induction hypothesis  $S_{k-1} \subseteq \Omega$  holds for this case since  $S_0 = \emptyset$ .

By line 2 of Algorithm 1, to show  $s^k \in \Omega$ , it is equivalent to show that

$$\max_{i \in \Omega} |\langle r^{k-1}, A_i \rangle| > \max_{j \in \Omega^c} |\langle r^{k-1}, A_j \rangle|. \quad (31)$$

In the following, we simplify (31). Since the minimum eigenvalue of  $A_{S_{k-1}}^T A_{S_{k-1}}$  is lower bounded by

$$1 - \delta_{|S_{k-1}|} \geq 1 - \delta_{|\Omega|} > 0,$$

$A_{S_{k-1}}^T A_{S_{k-1}}$  is invertible. Thus, by line 4 of Algorithm 1, we have

$$\hat{x}_{S_{k-1}} = (A_{S_{k-1}}^T A_{S_{k-1}})^{-1} A_{S_{k-1}}^T y. \quad (32)$$

Then, it follows from line 5 of Algorithm 1 and (32) that

$$\begin{aligned}
\mathbf{r}^{k-1} &= \mathbf{y} - \mathbf{A}_{S_{k-1}} \hat{\mathbf{x}}_{S_{k-1}} \\
&= (\mathbf{I} - \mathbf{A}_{S_{k-1}} (\mathbf{A}_{S_{k-1}}^T \mathbf{A}_{S_{k-1}})^{-1} \mathbf{A}_{S_{k-1}}^T) \mathbf{y} \\
&\stackrel{(a)}{=} \mathbf{P}_{S_{k-1}}^\perp (\mathbf{A} \mathbf{x} + \mathbf{v}) \\
&\stackrel{(b)}{=} \mathbf{P}_{S_{k-1}}^\perp (\mathbf{A}_\Omega \mathbf{x}_\Omega + \mathbf{v}) \\
&\stackrel{(c)}{=} \mathbf{P}_{S_{k-1}}^\perp (\mathbf{A}_{S_{k-1}} \mathbf{x}_{S_{k-1}} + \mathbf{A}_{\Omega \setminus S_{k-1}} \mathbf{x}_{\Omega \setminus S_{k-1}} + \mathbf{v}) \\
&\stackrel{(d)}{=} \mathbf{P}_{S_{k-1}}^\perp \mathbf{A}_{\Omega \setminus S_{k-1}} \mathbf{x}_{\Omega \setminus S_{k-1}} + \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}, \tag{33}
\end{aligned}$$

where (a) follows from the definition of  $\mathbf{P}_{S_{k-1}}^\perp$ , (b) is due to the fact that  $\Omega = \text{supp}(\mathbf{x})$ , (c) is from the induction assumption that  $S_{k-1} \subseteq \Omega$ , and (d) follows from

$$\mathbf{P}_{S_{k-1}}^\perp \mathbf{A}_{S_{k-1}} = \mathbf{0}. \tag{34}$$

Thus, by (33) and (34), for  $i \in S_{k-1}$ , we have

$$\langle \mathbf{r}^{k-1}, \mathbf{A}_i \rangle = \mathbf{A}_i^T \mathbf{r}^{k-1} = 0.$$

Therefore, to show (31), it is equivalent to show

$$\max_{i \in \Omega \setminus S_{k-1}} |\langle \mathbf{r}^{k-1}, \mathbf{A}_i \rangle| > \max_{j \in \Omega^c} |\langle \mathbf{r}^{k-1}, \mathbf{A}_j \rangle|. \tag{35}$$

In the following, we will use (33) to rewrite (35). By (33) and the (reverse) triangle inequality, we obtain

$$\begin{aligned}
&\max_{i \in \Omega \setminus S_{k-1}} |\langle \mathbf{r}^{k-1}, \mathbf{A}_i \rangle| \\
&= \|\mathbf{A}_{\Omega \setminus S_{k-1}}^T (\mathbf{P}_{S_{k-1}}^\perp \mathbf{A}_{\Omega \setminus S_{k-1}} \mathbf{x}_{\Omega \setminus S_{k-1}} + \mathbf{P}_{S_{k-1}}^\perp \mathbf{v})\|_\infty \\
&\geq \|\mathbf{A}_{\Omega \setminus S_{k-1}}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{A}_{\Omega \setminus S_{k-1}} \mathbf{x}_{\Omega \setminus S_{k-1}}\|_\infty \\
&\quad - \|\mathbf{A}_{\Omega \setminus S_{k-1}}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty, \tag{36}
\end{aligned}$$

and

$$\begin{aligned}
&\max_{j \in \Omega^c} |\langle \mathbf{r}^{k-1}, \mathbf{A}_j \rangle| \\
&= \|\mathbf{A}_{\Omega^c}^T (\mathbf{P}_{S_{k-1}}^\perp \mathbf{A}_{\Omega \setminus S_{k-1}} \mathbf{x}_{\Omega \setminus S_{k-1}} + \mathbf{P}_{S_{k-1}}^\perp \mathbf{v})\|_\infty \\
&\leq \|\mathbf{A}_{\Omega^c}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{A}_{\Omega \setminus S_{k-1}} \mathbf{x}_{\Omega \setminus S_{k-1}}\|_\infty + \|\mathbf{A}_{\Omega^c}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty. \tag{37}
\end{aligned}$$

Therefore, by (36) and (37), to show (35), it suffices to show

$$\begin{aligned}
&\|\mathbf{A}_{\Omega \setminus S_{k-1}}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{A}_{\Omega \setminus S_{k-1}} \mathbf{x}_{\Omega \setminus S_{k-1}}\|_\infty \\
&\quad - \|\mathbf{A}_{\Omega^c}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{A}_{\Omega \setminus S_{k-1}} \mathbf{x}_{\Omega \setminus S_{k-1}}\|_\infty \\
&> \|\mathbf{A}_{\Omega \setminus S_{k-1}}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty + \|\mathbf{A}_{\Omega^c}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty. \tag{38}
\end{aligned}$$

We next give a lower bound on the left-hand side of (38). By the induction assumption  $S_{k-1} \subseteq \Omega$ , we have

$$|\text{supp}(\mathbf{x}_{\Omega \setminus S_{k-1}})| = |\Omega| + 1 - k. \tag{39}$$

Thus,

$$\begin{aligned}
\|\mathbf{x}_{\Omega \setminus S_{k-1}}\|_2 &\geq \sqrt{|\Omega| + 1 - k} \min_{i \in \Omega \setminus S_{k-1}} |x_i| \\
&\geq \sqrt{|\Omega| + 1 - k} \min_{i \in \Omega} |x_i|. \tag{40}
\end{aligned}$$

Since  $S_{k-1} \subseteq \Omega$  and  $|S_{k-1}| = k - 1$ , by Lemma 1, we have

$$\begin{aligned}
&\|\mathbf{A}_{\Omega \setminus S_{k-1}}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{A}_{\Omega \setminus S_{k-1}} \mathbf{x}_{\Omega \setminus S_{k-1}}\|_\infty \\
&\quad - \|\mathbf{A}_{\Omega^c}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{A}_{\Omega \setminus S_{k-1}} \mathbf{x}_{\Omega \setminus S_{k-1}}\|_\infty \\
&\geq \frac{(1 - \sqrt{|\Omega| - k + 2\delta_{|\Omega|+1}}) \|\mathbf{x}_{\Omega \setminus S_{k-1}}\|_2}{\sqrt{|\Omega| + 1 - k}} \\
&\stackrel{(a)}{\geq} \frac{(1 - \sqrt{K + 1\delta_{|\Omega|+1}}) \|\mathbf{x}_{\Omega \setminus S_{k-1}}\|_2}{\sqrt{|\Omega| + 1 - k}} \\
&\stackrel{(b)}{\geq} \frac{(1 - \sqrt{K + 1\delta_{K+1}}) \|\mathbf{x}_{\Omega \setminus S_{k-1}}\|_2}{\sqrt{|\Omega| + 1 - k}} \\
&\stackrel{(c)}{\geq} (1 - \sqrt{K + 1\delta_{K+1}}) \min_{i \in \Omega} |x_i|, \tag{41}
\end{aligned}$$

where (a) is because  $k \geq 1$  and  $\mathbf{x}$  is  $K$ -sparse (i.e.,  $|\Omega| \leq K$ ), (b) follows from Lemma 2, and (c) is from (7) and (40).

We next give an upper bound on the right-hand side of (38). Obviously, there exist  $i_0 \in \Omega \setminus S_{k-1}$  and  $j_0 \in \Omega^c$  such that

$$\|\mathbf{A}_{\Omega \setminus S_{k-1}}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty = |\mathbf{A}_{i_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}|, \tag{42}$$

$$\|\mathbf{A}_{\Omega^c}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty = |\mathbf{A}_{j_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}|. \tag{43}$$

Hence,

$$\begin{aligned}
&\|\mathbf{A}_{\Omega \setminus S_{k-1}}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty + \|\mathbf{A}_{\Omega^c}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty \\
&= |\mathbf{A}_{i_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}| + |\mathbf{A}_{j_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}| \\
&= \|\mathbf{A}_{i_0 \cup j_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_1 \\
&\stackrel{(a)}{\leq} \sqrt{2} \|\mathbf{A}_{i_0 \cup j_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_2 \\
&\stackrel{(b)}{\leq} \sqrt{2(1 + \delta_{K+1})} \|\mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_2 \\
&\stackrel{(c)}{\leq} \sqrt{2(1 + \delta_{K+1})} \epsilon, \tag{44}
\end{aligned}$$

where (a) is due to that  $\mathbf{A}_{i_0 \cup j_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}$  is a  $2 \times 1$  vector and that  $\|\mathbf{x}\|_1 \leq \sqrt{|\text{supp}(\mathbf{x})|} \|\mathbf{x}\|_2$  for all  $\mathbf{x} \in \mathbb{R}^n$  (For more details, see e.g., [29, p.517]. Note that this inequality itself can be derived from the Cauchy-Schwarz inequality), (b) follows from Lemma 3 and (c) is because

$$\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_2 \leq \|\mathbf{v}\|_2 \leq \epsilon. \tag{45}$$

From (41) and (44), (38) (or equivalently (35)) can be guaranteed by

$$(1 - \sqrt{K + 1\delta_{K+1}}) \min_{i \in \Omega} |x_i| > \sqrt{2(1 + \delta_{K+1})} \epsilon,$$

i.e.,

$$\min_{i \in \Omega} |x_i| > \frac{\sqrt{2(1 + \delta_{K+1})} \epsilon}{1 - \sqrt{K + 1\delta_{K+1}}}.$$

Furthermore, by (7), we have  $\sqrt{1 + \delta_{K+1}} < \sqrt{2}$ . Thus, if (8) holds, then the OMP algorithm selects a correct index in each iteration.

Now, what remains to show is that the OMP algorithm performs exact  $|\Omega|$  iterations, which is equivalent to show that  $\|\mathbf{r}^k\|_2 > \epsilon$  for  $1 \leq k < |\Omega|$  and  $\|\mathbf{r}^{|\Omega|}\|_2 \leq \epsilon$ .



Since OMP selects a correct index at each iteration under (8), by the (reverse) triangle inequality and (33), for  $1 \leq k < |\Omega|$ , we have

$$\begin{aligned} \|\mathbf{r}^k\|_2 &= \|\mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k} + \mathbf{P}_{S_k}^\perp \mathbf{v}\|_2 \\ &\geq \|\mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k}\|_2 - \|\mathbf{P}_{S_k}^\perp \mathbf{v}\|_2 \\ &\stackrel{(a)}{\geq} \|\mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k}\|_2 - \epsilon \\ &\stackrel{(b)}{\geq} \sqrt{1 - \delta_{|\Omega|}} \|\mathbf{x}_{\Omega \setminus S_k}\|_2 - \epsilon \\ &\stackrel{(c)}{\geq} \sqrt{1 - \delta_{K+1}} \sqrt{|\Omega| - k} \min_{i \in \Omega} |x_i| - \epsilon \\ &\geq \sqrt{1 - \delta_{K+1}} \min_{i \in \Omega} |x_i| - \epsilon, \end{aligned} \quad (46)$$

where (a) is from (45), (b) is from Lemma 4, and (c) follows from Lemma 2 and (40). Therefore, if

$$\min_{i \in \Omega} |x_i| > \frac{2\epsilon}{\sqrt{1 - \delta_{K+1}}}, \quad (47)$$

then  $\|\mathbf{r}^k\|_2 > \epsilon$  for each  $1 \leq k < \Omega$ .

By some simple calculations, we can show that

$$\frac{2\epsilon}{1 - \sqrt{K+1}\delta_{K+1}} \geq \frac{2\epsilon}{\sqrt{1 - \delta_{K+1}}}. \quad (48)$$

Indeed, by the fact that  $0 < 1 - \delta_{K+1} < 1$ , we have

$$1 - \sqrt{K+1}\delta_{K+1} \leq 1 - \delta_{K+1} \leq \sqrt{1 - \delta_{K+1}}.$$

Thus, (48) holds.

Therefore, by (47) and (48), if (8) holds,  $\|\mathbf{r}^k\|_2 > \epsilon$  for each  $1 \leq k < \Omega$ , i.e., the OMP algorithm does not terminate before the  $|\Omega|$ -th iteration.

Similarly, by (33),

$$\begin{aligned} \|\mathbf{r}^{|\Omega|}\|_2 &= \|\mathbf{P}_{S_{|\Omega|}}^\perp \mathbf{A}_{\Omega \setminus S_{|\Omega|}} \mathbf{x}_{\Omega \setminus S_{|\Omega|}} + \mathbf{P}_{S_{|\Omega|}}^\perp \mathbf{v}\|_2 \\ &\stackrel{(a)}{=} \|\mathbf{P}_{S_{|\Omega|}}^\perp \mathbf{v}\|_2 \stackrel{(b)}{\leq} \epsilon, \end{aligned} \quad (49)$$

where (a) is because  $S_{|\Omega|} = |\Omega|$  and (b) follows from (45). So, by the stopping condition, the OMP algorithm terminates after the  $|\Omega|$ -th iteration. Therefore, the OMP algorithm performs  $|\Omega|$  iterations. This completes the proof.  $\square$

## APPENDIX C PROOF OF THEOREM 2

*Proof.* Similar to the proof of Theorem 1, we need to prove that the OMP algorithm selects correct indexes at all iterations and it performs exactly  $|\Omega|$  iterations.

We first prove the first part. By the proof of Theorem 1, we only need to prove that (38) holds. As the noise vector satisfies a different constraint, we need to give a new upper bound on the right-hand side of (38). To do this, we first use the method used in the proof of [8, Theorem 5] to give an upper bound on  $\|\mathbf{P}_{S_{k-1}} \mathbf{v}\|_2$  and then use (42) and (43) to given an upper bound on

$$\|\mathbf{A}_{\Omega \setminus S_{k-1}}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty + \|\mathbf{A}_{\Omega^c}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty.$$

Let  $\lambda$  denote the largest singular value of  $(\mathbf{A}_{S_{k-1}}^T \mathbf{A}_{S_{k-1}})^{-1}$ . Then  $\lambda$  equals to the reciprocal of the smallest singular value

of  $\mathbf{A}_{S_{k-1}}^T \mathbf{A}_{S_{k-1}}$ . Since  $\mathbf{A}$  satisfies the RIP of order  $K+1$  with  $\delta_{K+1}$ , the smallest singular value of  $\mathbf{A}_{S_{k-1}}^T \mathbf{A}_{S_{k-1}}$  cannot be smaller than  $1 - \delta_{K+1}$ . Thus,  $\lambda \leq 1/(1 - \delta_{K+1})$ . Therefore,

$$\begin{aligned} \|\mathbf{P}_{S_{k-1}} \mathbf{v}\|_2^2 &= \mathbf{v}^T \mathbf{P}_{S_{k-1}}^T \mathbf{P}_{S_{k-1}} \mathbf{v} = \mathbf{v}^T \mathbf{P}_{S_{k-1}} \mathbf{v} \\ &\stackrel{(a)}{=} \mathbf{v}^T \mathbf{A}_{S_{k-1}} (\mathbf{A}_{S_{k-1}}^T \mathbf{A}_{S_{k-1}})^{-1} \mathbf{A}_{S_{k-1}}^T \mathbf{v} \\ &\stackrel{(b)}{\leq} \lambda \|\mathbf{A}_{S_{k-1}}^T \mathbf{v}\|_2^2 \\ &\leq \frac{1}{1 - \delta_{K+1}} \|\mathbf{A}_{S_{k-1}}^T \mathbf{v}\|_2^2 \\ &\stackrel{(c)}{\leq} \frac{K-1}{1 - \delta_{K+1}} \|\mathbf{A}_{S_{k-1}}^T \mathbf{v}\|_\infty^2 \\ &\stackrel{(d)}{\leq} \frac{(K-1)\epsilon^2}{1 - \delta_{K+1}} < \frac{K\epsilon^2}{1 - \delta_{K+1}}, \end{aligned} \quad (50)$$

where (a) follows from the definition of  $\mathbf{P}_{S_{k-1}}$ , (b) is from the assumption that  $\lambda$  is the largest singular value of  $(\mathbf{A}_{S_{k-1}}^T \mathbf{A}_{S_{k-1}})^{-1}$ , (c) is because  $|S_{k-1}| = k-1 \leq K-1$ , and (d) follows from  $\|\mathbf{A}^T \mathbf{v}\|_\infty \leq \epsilon$ .

By (42), (43) and the triangular inequality, we have

$$\begin{aligned} &\|\mathbf{A}_{\Omega \setminus S_{k-1}}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty + \|\mathbf{A}_{\Omega^c}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_\infty \\ &= |\mathbf{A}_{i_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}| + |\mathbf{A}_{j_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}| \\ &= \|\mathbf{A}_{i_0 \cup j_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_1 \\ &\stackrel{(a)}{\leq} \sqrt{2} \|\mathbf{A}_{i_0 \cup j_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}\|_2 \\ &= \sqrt{2} \|\mathbf{A}_{i_0 \cup j_0}^T (\mathbf{I} - \mathbf{P}_{S_{k-1}}) \mathbf{v}\|_2 \\ &\leq \sqrt{2} \|\mathbf{A}_{i_0 \cup j_0}^T \mathbf{v}\|_2 + \sqrt{2} \|\mathbf{A}_{i_0 \cup j_0}^T \mathbf{P}_{S_{k-1}} \mathbf{v}\|_2 \\ &\stackrel{(b)}{\leq} 2 \|\mathbf{A}_{i_0 \cup j_0}^T \mathbf{v}\|_\infty + \sqrt{2(1 + \delta_2)} \|\mathbf{P}_{S_{k-1}} \mathbf{v}\|_2 \\ &\stackrel{(c)}{\leq} 2\epsilon + \sqrt{\frac{1 + \delta_2}{1 - \delta_{K+1}}} \sqrt{2K} \epsilon, \end{aligned} \quad (51)$$

where (a) is due to the fact that  $\mathbf{A}_{i_0 \cup j_0}^T \mathbf{P}_{S_{k-1}}^\perp \mathbf{v}$  is a  $2 \times 1$  vector and the Cauchy-Schwarz inequality, (b) and (c) respectively follow from Lemma 3 and (50).

Therefore, by (41) and (51), if (13) holds, then (38) holds,<sup>5</sup> i.e., OMP selects correct indexes in all iterations if (13) holds.

Our next job is to prove that the OMP algorithm does not terminate before the  $|\Omega|$ -th iteration. By the (reverse) triangular inequality and (33), we have

$$\begin{aligned} \|\mathbf{A}^T \mathbf{r}^k\|_\infty &= \|\mathbf{A}^T (\mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k} + \mathbf{P}_{S_k}^\perp \mathbf{v})\|_\infty \\ &\geq \|\mathbf{A}_{\Omega \setminus S_k}^T (\mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k} + \mathbf{P}_{S_k}^\perp \mathbf{v})\|_\infty \\ &\geq \|\mathbf{A}_{\Omega \setminus S_k}^T \mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k}\|_\infty - \|\mathbf{A}_{\Omega \setminus S_k}^T \mathbf{P}_{S_k}^\perp \mathbf{v}\|_\infty. \end{aligned} \quad (52)$$

In the following, we give a lower bound on

$$\|\mathbf{A}_{\Omega \setminus S_k}^T \mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k}\|_\infty.$$

<sup>5</sup>Note that (38) still holds under the relaxed condition of (13).

It is not hard to check that

$$\begin{aligned}
& \| \mathbf{A}_{\Omega \setminus S_k}^T \mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k} \|_\infty \\
& \geq \frac{1}{\sqrt{|\Omega| - k}} \| \mathbf{A}_{\Omega \setminus S_k}^T \mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k} \|_2 \\
& = \frac{\| \mathbf{x}_{\Omega \setminus S_k} \|_2 \| \mathbf{A}_{\Omega \setminus S_k}^T \mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k} \|_2}{\sqrt{|\Omega| - k} \| \mathbf{x}_{\Omega \setminus S_k} \|_2} \\
& \stackrel{(a)}{\geq} \frac{|\mathbf{x}_{\Omega \setminus S_k}^T \mathbf{A}_{\Omega \setminus S_k}^T \mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k}|}{\sqrt{|\Omega| - k} \| \mathbf{x}_{\Omega \setminus S_k} \|_2} \\
& \stackrel{(b)}{=} \frac{\| \mathbf{P}_{S_k}^\perp \mathbf{A}_{\Omega \setminus S_k} \mathbf{x}_{\Omega \setminus S_k} \|_2^2}{\sqrt{|\Omega| - k} \| \mathbf{x}_{\Omega \setminus S_k} \|_2} \\
& \stackrel{(c)}{\geq} \frac{(1 - \delta_K) \| \mathbf{x}_{\Omega \setminus S_k} \|_2^2}{\sqrt{|\Omega| - k} \| \mathbf{x}_{\Omega \setminus S_k} \|_2} \\
& \geq (1 - \delta_{K+1}) \frac{\| \mathbf{x}_{\Omega \setminus S_k} \|_2}{\sqrt{|\Omega| - k}} \\
& \geq (1 - \delta_{K+1}) \min_{i \in \Omega} |x_i|,
\end{aligned} \tag{53}$$

where (a) follows from the Cauchy-Schwarz inequality, (b) is due to (19), (c) is from Lemma 4, and the last inequality is from (40).

In the following, we give an upper bound on

$$\| \mathbf{A}_{\Omega \setminus S_k}^T \mathbf{P}_{S_k}^\perp \mathbf{v} \|_\infty.$$

Let  $j_0 \in \Omega \setminus S_k$  such that

$$\| \mathbf{A}_{\Omega \setminus S_k}^T \mathbf{P}_{S_k}^\perp \mathbf{v} \|_\infty = | \mathbf{A}_{j_0}^T \mathbf{P}_{S_k}^\perp \mathbf{v} |.$$

Then, by the triangular inequality, we obtain

$$\begin{aligned}
& \| \mathbf{A}_{\Omega \setminus S_k}^T \mathbf{P}_{S_k}^\perp \mathbf{v} \|_\infty = | \mathbf{A}_{j_0}^T (\mathbf{I} - \mathbf{P}_{S_k}) \mathbf{v} | \\
& \leq | \mathbf{A}_{j_0}^T \mathbf{v} | + | \mathbf{A}_{j_0}^T \mathbf{P}_{S_k} \mathbf{v} | \leq \epsilon + | \mathbf{A}_{j_0}^T \mathbf{P}_{S_k} \mathbf{v} | \\
& \leq \left( 1 + \sqrt{\frac{1 + \delta_1}{1 - \delta_{K+1}}} \sqrt{K} \right) \epsilon,
\end{aligned} \tag{54}$$

where the last inequality follows from the Cauchy-Schwarz inequality, (50) and Lemma 3.

Therefore, for each  $1 \leq k < |\Omega|$ , by (13) and (52)-(54), we have

$$\begin{aligned}
& \| \mathbf{A}^T \mathbf{r}^k \|_\infty \\
& \geq (1 - \delta_{K+1}) \min_{i \in \Omega} |x_i| - \left( 1 + \sqrt{\frac{1 + \delta_1}{1 - \delta_{K+1}}} \sqrt{K} \right) \epsilon \\
& > \frac{2(1 - \delta_{K+1})}{1 - \sqrt{K+1}\delta_{K+1}} \left( 1 + \sqrt{\frac{1 + \delta_2}{1 - \delta_{K+1}}} \sqrt{K} \right) \epsilon \\
& \quad - \left( 1 + \sqrt{\frac{1 + \delta_1}{1 - \delta_{K+1}}} \sqrt{K} \right) \epsilon \\
& \stackrel{(a)}{\geq} 2 \left( 1 + \sqrt{\frac{1 + \delta_2}{1 - \delta_{K+1}}} \sqrt{K} \right) \epsilon - \left( 1 + \sqrt{\frac{1 + \delta_1}{1 - \delta_{K+1}}} \sqrt{K} \right) \epsilon \\
& \geq \left( 1 + \sqrt{\frac{1 + \delta_2}{1 - \delta_{K+1}}} \sqrt{K} \right) \epsilon,
\end{aligned}$$

where (a) is because

$$1 - \delta_{K+1} \geq 1 - \sqrt{K+1}\delta_{K+1},$$

and the last inequality follows from Lemma 2. So, by the stopping condition (12), the OMP algorithm does not terminate before the  $|\Omega|$ -th iteration.<sup>6</sup>

Finally, we prove that OMP terminates after performing the  $|\Omega|$ -th iteration. By (49), we have

$$\mathbf{r}^{|\Omega|} = \mathbf{P}_{S_{|\Omega|}}^\perp \mathbf{v}.$$

Thus, applying some techniques which are similar to that for deriving (54), we obtain

$$\begin{aligned}
\| \mathbf{A}^T \mathbf{r}^{|\Omega|} \|_\infty & = \| \mathbf{A}^T \mathbf{P}_{S_{|\Omega|}}^\perp \mathbf{v} \|_\infty \\
& \leq \left( 1 + \sqrt{\frac{1 + \delta_1}{1 - \delta_{K+1}}} \sqrt{K} \right) \epsilon, \\
& \leq \left( 1 + \sqrt{\frac{1 + \delta_2}{1 - \delta_{K+1}}} \sqrt{K} \right) \epsilon,
\end{aligned}$$

By the stopping condition (12), the OMP algorithm terminates after performing the  $|\Omega|$ -th iteration.<sup>7</sup>  $\square$

#### APPENDIX D PROOF OF THEOREM 3

*Proof.* To prove the theorem, it suffices to show that there exists a linear model of the form (1), where  $\mathbf{v}$  satisfies  $\| \mathbf{v} \|_2 \leq \epsilon$ ,  $\mathbf{A}$  satisfies the RIP with  $\delta_{K+1} = \delta$  for any given  $\delta$  satisfying (15), and  $\mathbf{x}$  is  $K$ -sparse and satisfies

$$\min_{i \in \Omega} |x_i| = \gamma \tag{55}$$

for some  $\gamma$  satisfying

$$0 < \gamma < \frac{\sqrt{1 - \delta} \epsilon}{1 - \sqrt{K+1}\delta}, \tag{56}$$

such that the OMP algorithm fails to recover the support of  $\mathbf{x}$  in  $K$  iterations.

In the following, we construct such a linear model.

Let  $\mathbf{1}_K$  be a  $K$ -dimensional column vector with all entries being 1, then there exist  $\xi_i \in \mathbb{R}^K$ ,  $1 \leq i \leq K-1$ , such that

$$\begin{bmatrix} \xi_1 & \xi_2 & \cdots & \xi_{K-1} & \frac{1}{\sqrt{K}} \mathbf{1}_K \end{bmatrix} \in \mathbb{R}^{K \times K}$$

is an orthogonal matrix. Let

$$\mathbf{U} = \begin{bmatrix} \xi_1 & \xi_2 & \cdots & \xi_{K-1} & \frac{\mathbf{1}_K}{\sqrt{K(\beta^2+1)}} & \frac{\beta \mathbf{1}_K}{\sqrt{K(\beta^2+1)}} \\ 0 & 0 & \cdots & 0 & \frac{\beta}{\sqrt{\beta^2+1}} & -\frac{1}{\sqrt{\beta^2+1}} \end{bmatrix}^T, \tag{57}$$

<sup>6</sup>If  $\mathbf{A}$  is column normalized, then  $\delta_1 = 0$ . Thus, under the relaxed condition of (13), we have  $\| \mathbf{A}^T \mathbf{r}^k \|_\infty \geq \left( 1 + \frac{\sqrt{K}}{\sqrt{1 - \delta_{K+1}}} \right) \epsilon$ . Thus, the OMP algorithm does not terminate before the  $|\Omega|$ -th iteration under the relaxed stopping condition.

<sup>7</sup>If  $\mathbf{A}$  is column normalized, then  $\delta_1 = 0$ . Thus,  $\| \mathbf{A}^T \mathbf{r}^{|\Omega|} \|_\infty \leq \left( 1 + \frac{\sqrt{K}}{\sqrt{1 - \delta_{K+1}}} \right) \epsilon$ . Thus, the OMP algorithm terminates after performing the  $|\Omega|$ -th iteration.

where

$$\beta = \frac{\sqrt{K+1} - 1}{\sqrt{K}}.$$

Then, it is not hard to prove that  $\mathbf{U}$  is also an orthogonal matrix. Applying some techniques which are similar to that for deriving (22), we can show that

$$\frac{1 - \beta^2}{1 + \beta^2} = \frac{1}{\sqrt{K+1}}, \quad \frac{2\beta}{1 + \beta^2} = \frac{\sqrt{K}}{\sqrt{K+1}}. \quad (58)$$

Let  $\mathbf{D} \in \mathbb{R}^{(K+1) \times (K+1)}$  be the diagonal matrix with

$$d_{ii} = \begin{cases} \sqrt{1 - \delta} & i = K \\ \sqrt{1 + \delta} & i \neq K \end{cases}, \quad (59)$$

and

$$\mathbf{A} = \mathbf{D}\mathbf{U}, \quad (60)$$

then  $\mathbf{A}^T \mathbf{A} = \mathbf{U}^T \mathbf{D}^2 \mathbf{U}$ . It is not hard to see that  $\mathbf{A}$  satisfies the RIP with  $\delta_{K+1} = \delta$ . In fact, let  $\mathbf{V} = \mathbf{U}^T$ , then by the fact that  $\mathbf{U}$  is orthogonal, we have

$$\mathbf{A}^T \mathbf{A} = \mathbf{U}^T \mathbf{D}^2 \mathbf{U} = \mathbf{V} \mathbf{D}^2 \mathbf{V}^T = \mathbf{V} \mathbf{D}^2 \mathbf{V}^{-1}.$$

Thus,  $\mathbf{V} \mathbf{D}^2 \mathbf{V}^{-1}$  is a valid eigenvalue decomposition of  $\mathbf{A}^T \mathbf{A}$ , and  $\mathbf{D}^2$  is the diagonal matrix containing the eigenvalues. Therefore, the largest and smallest eigenvalues of  $\mathbf{A}^T \mathbf{A}$  are respectively  $1 + \delta$  and  $1 - \delta$ , which implies that  $\delta_{K+1} = \delta$ .

Let

$$\mathbf{x} = [\gamma \mathbf{1}_K^T \quad 0]^T \in \mathbb{R}^{K+1} \quad (61)$$

for any  $\gamma$  satisfying (56) (recall that  $\mathbf{1}_K$  is a  $K$ -dimensional column vector with all of its entries being 1). Then  $\mathbf{x}$  is  $K$ -sparse and satisfies (55). By (57) and the fact that  $\xi_i^T \mathbf{1}_K = 0$  for  $1 \leq i \leq K - 1$ , we have

$$\mathbf{U}\mathbf{x} = \sqrt{\frac{K}{\beta^2 + 1}} \gamma [\mathbf{0}_{K-1}^T \quad 1 \quad \beta]^T.$$

Thus,

$$\mathbf{D}^2 \mathbf{U}\mathbf{x} = \sqrt{\frac{K}{\beta^2 + 1}} \gamma [\mathbf{0}_{K-1}^T \quad 1 - \delta \quad (1 + \delta)\beta]^T. \quad (62)$$

Let

$$\mathbf{v} = \mathbf{D}^{-1} \mathbf{U} [\mathbf{0}_K^T \quad -\sqrt{1 - \delta}\epsilon]^T \in \mathbb{R}^{K+1}, \quad (63)$$

then, by (60) and the fact that  $\mathbf{U}$  is orthogonal, we have

$$\mathbf{A}^T \mathbf{v} = [\mathbf{0}_K^T \quad -\sqrt{1 - \delta}\epsilon]^T. \quad (64)$$

In the following, we show that  $\|\mathbf{v}\|_2 \leq \epsilon$ . By (57) and (63), we have

$$\mathbf{v} = \mathbf{D}^{-1} \frac{\epsilon}{\sqrt{\beta^2 + 1}} [\mathbf{0}_{K-1}^T \quad -\beta\sqrt{1 - \delta} \quad \sqrt{1 - \delta}]^T.$$

Thus, by (59), we obtain

$$\mathbf{v} = \frac{\epsilon}{\sqrt{\beta^2 + 1}} [\mathbf{0}_{K-1}^T \quad -\beta \quad \sqrt{\frac{1 - \delta}{1 + \delta}}]^T,$$

and hence  $\|\mathbf{v}\|_2 \leq \epsilon$ .

Let  $\mathbf{e}_i, 1 \leq i \leq K$ , be the  $i$ -th column of the  $(K + 1) \times (K + 1)$  identity matrix, then, for  $1 \leq i \leq K$ , we have

$$\begin{aligned} \langle \mathbf{y}, \mathbf{A}_i \rangle &= \langle \mathbf{A}\mathbf{x} + \mathbf{v}, \mathbf{A}_i \rangle = \mathbf{e}_i^T \mathbf{A}^T \mathbf{A}\mathbf{x} + \mathbf{e}_i^T \mathbf{A}^T \mathbf{v} \\ &\stackrel{(a)}{=} \mathbf{e}_i^T \mathbf{U}^T \mathbf{D}^2 \mathbf{U}\mathbf{x} \\ &\stackrel{(b)}{=} \left( \frac{(1 - \delta) + (1 + \delta)\beta^2}{\beta^2 + 1} \right) \gamma \\ &= \left( 1 - \frac{1 - \beta^2}{1 + \beta^2} \delta \right) \gamma \\ &= \left( 1 - \frac{1}{\sqrt{K+1}} \delta \right) \gamma, \end{aligned}$$

where (a) follows from the fact that  $\mathbf{U}$  is orthogonal, (60) and (64), (b) is due to (57) and (62), and the last equality is from (58). Thus, by (15),

$$\max_{i \in \Omega} |\langle \mathbf{y}, \mathbf{A}_i \rangle| = \left( 1 - \frac{1}{\sqrt{K+1}} \delta \right) \gamma. \quad (65)$$

Similarly, we have

$$\begin{aligned} \langle \mathbf{y}, \mathbf{A}_{K+1} \rangle &= \mathbf{e}_{K+1}^T \mathbf{U}^T \mathbf{D}^2 \mathbf{U}\mathbf{x} + \mathbf{e}_{K+1}^T \mathbf{A}^T \mathbf{v} \\ &= \frac{-2\beta}{\beta^2 + 1} \sqrt{K} \delta \gamma - \sqrt{1 - \delta} \epsilon \\ &= \frac{-K}{\sqrt{K+1}} \delta \gamma - \sqrt{1 - \delta} \epsilon, \end{aligned}$$

where the last equality is from (58). Thus

$$\max_{j \in \Omega^c} |\langle \mathbf{y}, \mathbf{A}_j \rangle| = \frac{K}{\sqrt{K+1}} \delta \gamma + \sqrt{1 - \delta} \epsilon. \quad (66)$$

By (56), (65) and (66), we have

$$\max_{i \in \Omega} |\langle \mathbf{y}, \mathbf{A}_i \rangle| < \max_{j \in \Omega^c} |\langle \mathbf{y}, \mathbf{A}_j \rangle|.$$

Thus, by line 2 of Algorithm 1, OMP chooses an index in  $\Omega^c$  in the first iteration. Therefore, OMP fails to recover  $\Omega$  in  $K$  iterations. This completes the proof.  $\square$

## APPENDIX E PROOF OF THEOREM 4

*Proof.* Similar to the proof of Theorem 3, we need to show that there exists a linear model of the form (1), where  $\mathbf{v}$  satisfies  $\|\mathbf{A}^T \mathbf{v}\|_\infty \leq \epsilon$ ,  $\mathbf{A}$  satisfies the RIP with  $\delta_{K+1} = \delta$  for any given  $\delta$  satisfying (15), and  $\mathbf{x}$  is  $K$ -sparse and satisfies (55) with

$$0 < \gamma < \frac{2\epsilon}{1 - \sqrt{K+1}\delta}, \quad (67)$$

such that the OMP algorithm fails to recover the support of  $\mathbf{x}$  in  $K$  iterations.

We use the same  $\mathbf{A}$  and  $\mathbf{x}$  (see (60) and (61)) as in the proof of Theorem 3, but instead of (63), we define

$$\mathbf{v} = -\epsilon \mathbf{D}^{-1} \mathbf{U} \mathbf{1}_{K+1} \in \mathbb{R}^{K+1}. \quad (68)$$

Recall that  $\mathbf{1}_{K+1}$  is a  $(K+1)$ -dimensional column vector with all of its entries being 1. So by (60), we obtain,

$$\mathbf{A}^T \mathbf{v} = -\epsilon \mathbf{1}_n,$$

leading to  $\|\mathbf{A}^T \mathbf{v}\|_\infty = \epsilon$ .

Applying some techniques which are similar to that for deriving (65) and (66), we have

$$\max_{i \in \Omega} |\langle \mathbf{y}, \mathbf{A}_i \rangle| = \left(1 - \frac{1}{\sqrt{K+1}}\delta\right) \gamma - \epsilon.$$

and

$$\max_{j \in \Omega^c} |\langle \mathbf{y}, \mathbf{A}_j \rangle| = \frac{K}{\sqrt{K+1}}\delta\gamma + \epsilon.$$

Thus, by (67), we further have

$$\max_{i \in \Omega} |\langle \mathbf{y}, \mathbf{A}_i \rangle| < \max_{j \in \Omega^c} |\langle \mathbf{y}, \mathbf{A}_j \rangle|.$$

By line 2 of Algorithm 1, OMP chooses an index in  $\Omega^c$  in the first iteration. Therefore, OMP fails to recover  $\Omega$  in  $K$  iterations. This completes the proof.  $\square$

## REFERENCES

- [1] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best  $k$ -term approximation," *J. Amer. Math. Soc.*, vol. 22, pp. 211–231, 2009.
- [4] J. Wen, D. Li, and F. Zhu, "Stable recovery of sparse signals via  $l_p$ -minimization," *Appl. Comput. Harmon. Anal.*, vol. 38, no. 1, pp. 161–176, 2015.
- [5] J. J. Fuchs., "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3601–3608, 2005.
- [6] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, pp. 6–18, 2006.
- [7] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, no. 11, pp. 589–592, 2008.
- [8] T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, pp. 4680–4688, 2011.
- [9] E. J. Candès and T. Tao, "The dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ ," *Ann. Statist.*, vol. 35, pp. 2313–2351, 2007.
- [10] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [11] Q. Mo and S. Li, "New bounds on the restricted isometry constant  $\delta_{2k}$ ," *Appl. Comput. Harmon. Anal.*, vol. 31, pp. 460–468, 2011.
- [12] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals, Systems, and Computers*, vol. 1. IEEE, Pacific Grove, CA, Nov. 1993, pp. 40–44.
- [13] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [14] M. Davenport and M. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, 2010.
- [15] E. Liu and V. Temlyakov, "The orthogonal super greedy algorithm and applications in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2040–2047, 2012.
- [16] Q. Mo and S. Yi, "A remark on the restricted isometry property in orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3654–3656, 2012.
- [17] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4973–4976, 2012.
- [18] L.-H. Chang and J.-Y. Wu, "An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 707–710, 2014.
- [19] Q. Mo, "A sharp restricted isometry constant bound of orthogonal matching pursuit," *arXiv:1501.01708*, 2015.
- [20] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [21] J. Wen, X. Zhu, and D. Li, "Improved bounds on the restricted isometry constant for orthogonal matching pursuit," *Electronics Letters*, vol. 49, pp. 1487–1489, 2013.
- [22] Y. Shen and S. Li, "Sparse signals recovery from noisy measurements by orthogonal matching pursuit," *Inverse Problems and Imaging*, vol. 9, no. 1, pp. 231–238, 2015.
- [23] R. Wu, W. Huang, and D. Chen, "The exact support recovery of sparse signals with noise via orthogonal matching pursuit," *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 403–406, 2013.
- [24] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, "Exact recovery conditions for sparse representations with partial support information," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7509–7524, Nov 2013.
- [25] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, "A sharp condition for exact support recovery of sparse signals with orthogonal matching pursuit," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 2364–2368.
- [26] J. Wang, "Support recovery with orthogonal matching pursuit in the presence of noise," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5868–5877, Nov. 2015.
- [27] D. Needel and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
- [28] Y. Shen, B. Li, W. Pan, and J. Li, "Analysis of generalized orthogonal matching pursuit using restricted isometry constant," *Electron. Lett.*, vol. 50, no. 14, pp. 1020–1022, 2014.
- [29] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer-Verlag, 2013.